# Problem 1:

See the code Q1.py and helper.py for data formatting.

# Problem 2:

Accuracy is: 0.8269049858889934

tn: 133, tp: 746, fn: 83, fp: 101

# Problem 3:

Accuracy is: 0.8993414863593603

tn: 177, tp: 779, fn: 50, fp: 57

# Problem 4:

For the random forest, I’m using random state = 50 for a consistent output

A graph of different colored lines

Description automatically generated

Best N = 7, best d = 5

The highest accuracy: 0.9049858889934148

tn: 145, tp: 807, fn: 22, fp: 89

# Problem 5:

No code needed for this part.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | TP | TN | FP | FN | Accuracy | TPR | TNR |
| Naïve Bayesian | 746 | 133 | 101 | 83 | 0.8269 | 0.8999 | 0.5684 |
| Decision Tree | 779 | 177 | 57 | 50 | 0.8993 | 0.9367 | 0.7564 |
| Random Forest | 807 | 145 | 89 | 22 | 0.905 | 0.9735 | 0.6197 |

Random Forest yields the highest accuracy rate. However, it’s not much higher than decision tree’s accuracy, while the TNR is much lower than that of decision tree.

Both Decision tree and random forest are significantly better than Naïve Bayesian.